Utrecht University

DEMEVA WORKSHOP 2026 · UTRECHT · 16 OCTOBER

Gaze as Evidence for Common Grounding

A cross-corpus analysis of MapTask and MUNDEX

Nan Li · Albert Gatt · Massimo Poesio

Utrecht University · Queen Mary University of London

To appear at MINT @ EMNLP 2026 · arXiv:2609.18011

One exchange, two understandings



WHY

Each relies on what they take to be shared, their **common ground** , but on different parts of it. →

Meaning variation inside the conversation:

variation between interlocutors' interpretations

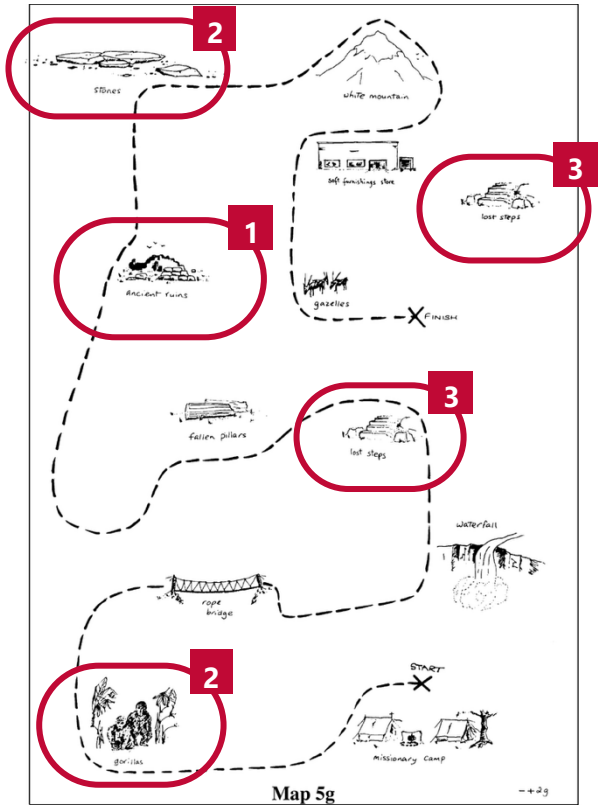
→

IN JOINT TASKS

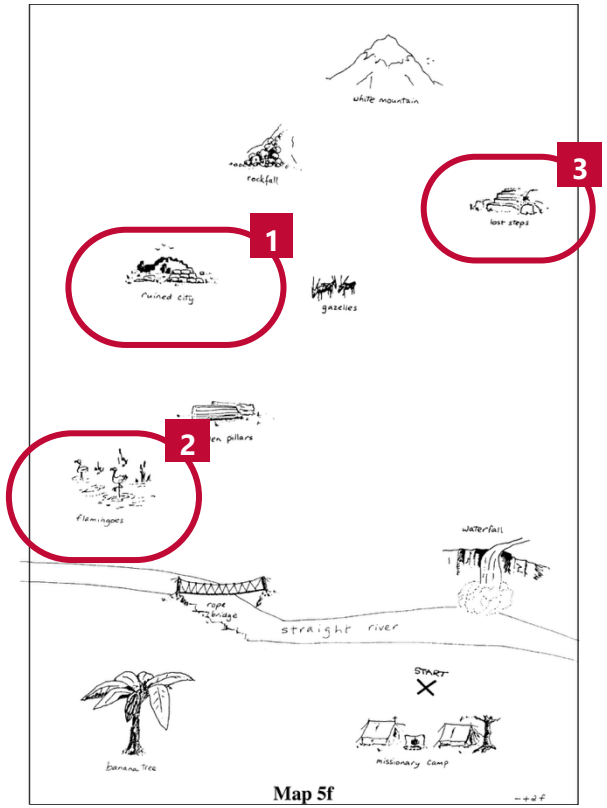
Asymmetric information makes it more likely. Some tasks build it in: one partner knows more, or the two have different information.

Common ground: what partners mutually take to be shared (Clark & Brennan 1991).

Two private maps that differ by design



GIVER · HAS THE ROUTE



FOLLOWER · DRAWS IT

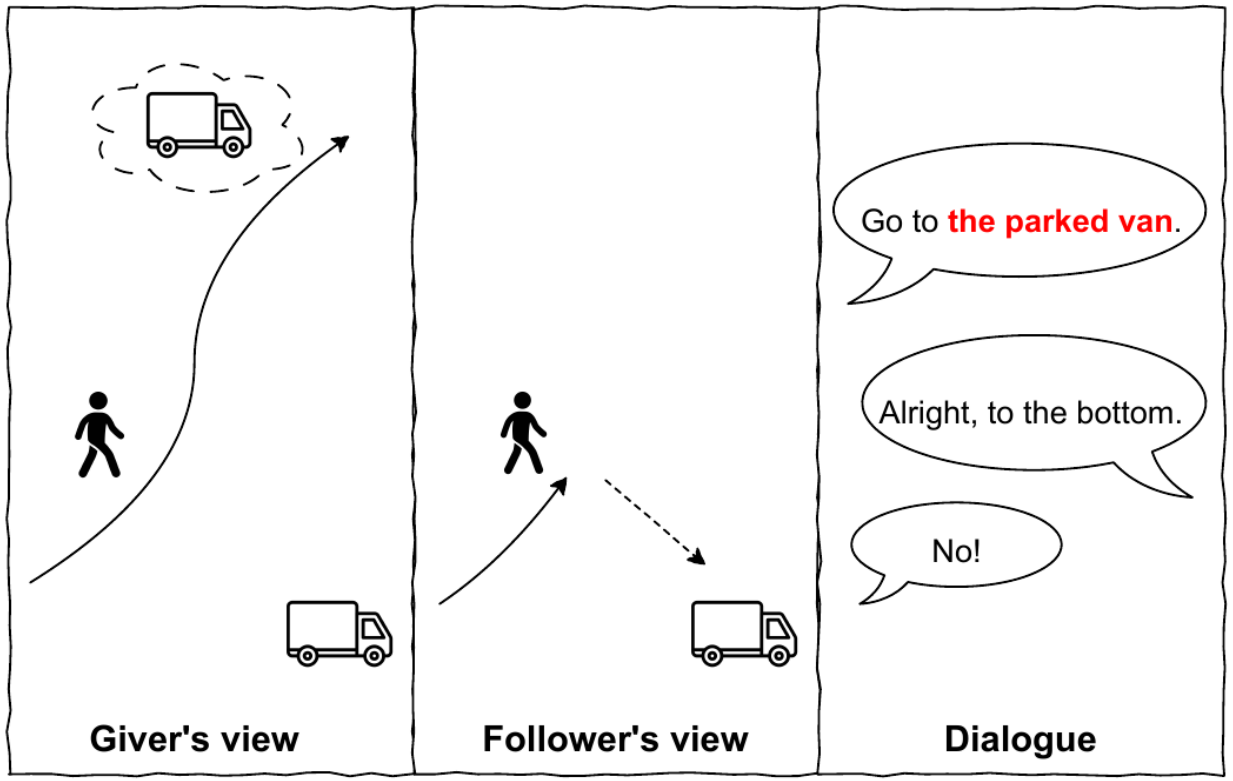
Three types of landmark discrepancies

1. **Lexical** — one place, two names: “ancient ruins” vs “ruined city”
2. **Existence** — on one map only: stones, gorillas vs flamingoes
3. **Multiplicity** — “lost steps” twice for the giver, once for the follower

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S., & Weinert, R. (1991). The HCRC Map Task Corpus. Language and Speech, 34(4), 351–366.

TASKS · MAPTASK LABELS

“Go to the parked van”: one expression, two interpretations



PERSPECTIVIST LABELS

Our LREC 2026 annotation (Li et al. 2026): for every reference, the speaker’s **intended** landmark and the addressee’s **interpreted** landmark, recorded separately.

Multiplicity , as here: 7% of references, about half of all misunderstandings.

ALIGNED same landmark

PENDING not yet resolved

MISUNDERSTOOD different referents

TASKS · MUNDEX LABELS

Two perspectives on one understanding event

EX (Explainer) teaches EE (Explainee) a board game; afterwards, each labels EE's understanding from the video.

ONE REAL EXCHANGE · TRANSLATED AND SHORTENED

EX At some point you have to say: okay, from now on I'm turning back. And then you can only go back.

EE Okay. Do I lose oxygen doing that too?

EX … If you have treasures, yes. If you have no treasures, no.

EE Yeah. … Not then, okay.

MUNDEX EX_13–EE_c, 03:25–03:37.

ASIDE · MEANING VARIATION

Across the corpus, **44 of 135** linked annotation pairs disagree on UND vs non-UND.



91 agree · 44 disagree

We keep both labels as separate observations.

Explainer · judges the EE — PARTIALLY UNDERSTOOD

"She understood the rule that she uses oxygen, but not quite when it applies."

Explainee · self-report — UNDERSTOOD

"I understood that well too, because it was phrased precisely."

Türk, O., Wagner, P., Buschmeier, H., Grimminger, A., Wang, Y., & Lazarov, S. (2023). MUNDEX: A multimodal corpus for the study of the understanding of explanations. Proceedings of the 1st International Multimodal Communication Symposium.

Grounding runs on evidence; gaze may supply some of it

THEORY · EVIDENCE OF UNDERSTANDING

roughly weaker → stronger (Clark & Schaefer 1989)

1. Continued attention
2. Initiation of the relevant next contribution
3. Acknowledgement
4. Demonstration
5. Display

OUR HYPOTHESES · WHAT GAZE MAY SHOW

TASK GAZE — continued attention to the joint task

PARTNER GAZE — attention to the partner; monitoring and inviting feedback (Bavelas et al. 2002; Clark & Krych 2004) ; can also signal trouble

AWAY GAZE — processing load or a topic shift (Doherty-Sneddon & Phelps 2005; Lazarov & Grimminger 2026)

Our categories record **where** people look: one look can serve several of these functions.

BACKGROUND · RELATED WORK

What earlier gaze studies found, and what we add

MAP TASKS · ROUTE GIVING

Boyle et al. 1994	more looks up around landmarks that differ between maps
Nakano et al. 2003	listener gaze at the speaker precedes elaboration; map gaze precedes the next instruction
Murat & Vogel 2026	partner gaze at turn ends accompanies difficulty

MUNDEX · EXPLANATIONS

Wang et al. 2026	gaze entropy (surprisal of automatically estimated gaze labels) and other cognitive-load cues, added to text features, improve prediction of self-reported understanding
Lazarov & Grimminger 2026	gaze aversion goes with topic shifts

THIS STUDY ADDS

Labels per reference (MapTask)

intended vs interpreted landmark

One vocabulary, two corpora

partner · task · away

Both perspectives

Giver vs. follower
EX judgment vs. EE self-report

RESEARCH QUESTIONS

Three questions about gaze and grounding

Q1 ASSOCIATIONS

Which gaze patterns go with grounding in both tasks?

Overall; then by who produced the reference, visibility, and whose label

MAPTASK MUNDEX

Q2 PROCESS

How does gaze change when a referent becomes aligned?

Same-speaker pairs along reference chains

MAPTASK

Q3 PREDICTION

How well do gaze features predict grounding labels, and which feature groups help?

Held-out dialogues and explainers; vs controls-only; feature-group ablation

MAPTASK MUNDEX

METHOD · DATA AND WINDOWS

Two asymmetric tasks, two kinds of grounding label

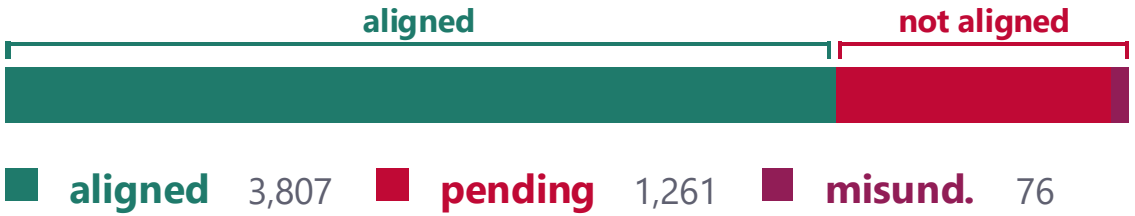
HCRC MAPTASK · ENGLISH

Route reproducing with**maps that differ**

46
dialogues
31 eye contact · 15 without

5,144
windows
reference + 1.5 s

74%
aligned



aligned 3,807 **pending** 1,261 **misund.** 76

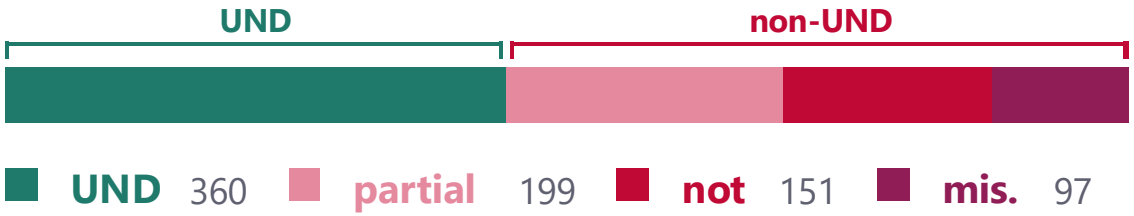
MUNDEX · GERMAN

Explaining a **board game** to a newcomer

26
interactions
9 explainers

807
windows
annotation ± 2 s

458 · 349
labelled
by explainer · by explainee



UND 360 **partial** 199 **not** 151 **mis.** 97

Windows with < 30% gaze coverage are dropped.

METHOD · GAZE FEATURES

One gaze vocabulary: partner · task · away

MAPTASK	MUNDEX		SHARED
up*	EX / EE	→	partner
down	TABLE	→	task
off	AWAY	→	away

* literally partner-directed only with eye contact

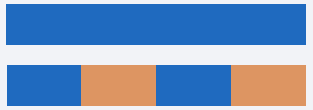
RAW PROPORTIONS · PER WINDOW

Share of time on each target, per person · mutual gaze

STRUCTURED = RAW + …

Entropy: how spread gaze time is

Transitions: how often gaze switches



entropy 0 · 0 switches

entropy 1 bit · 3 switches

Two 2-s windows: **task** · **partner**

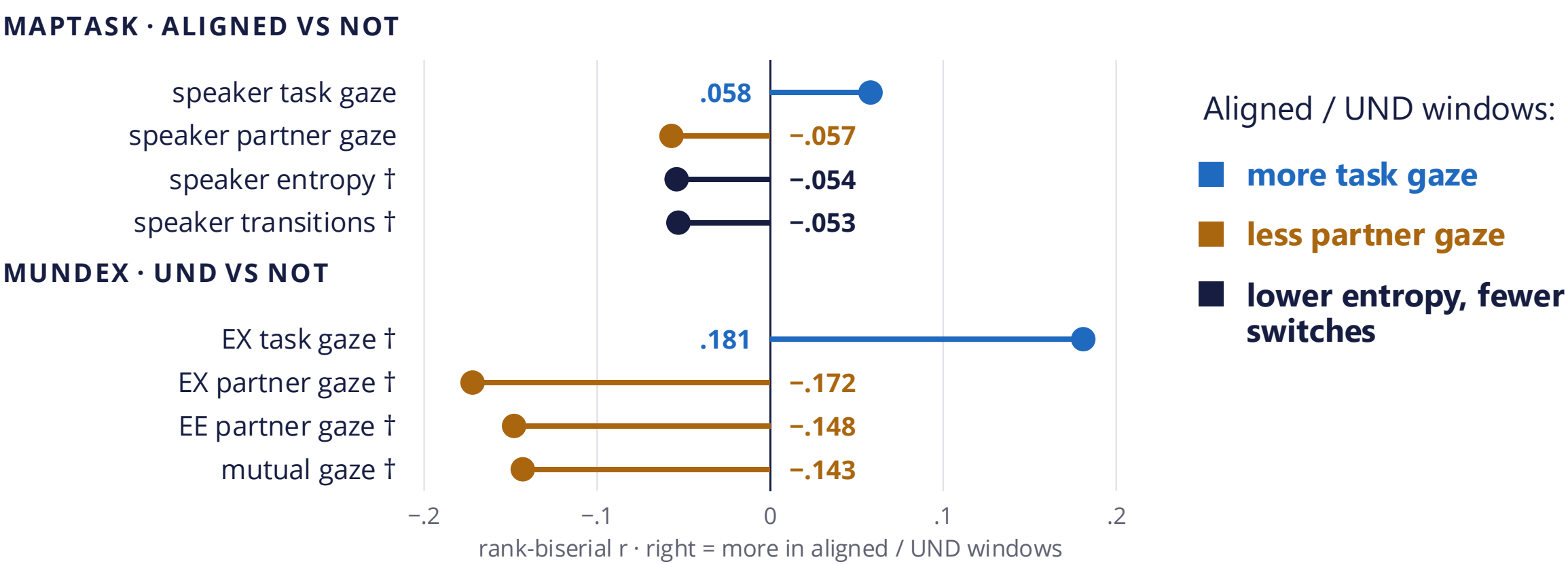
EXTENDED GROUPS · prediction only

temporal (timing) · bigrams (switch direction) · coordination (joint gaze) · ratios

Coded from video, not eye-tracked: the three-target vocabulary can map onto corpora with similar discrete gaze coding.

Q1 ASSOCIATIONS · OVERALL

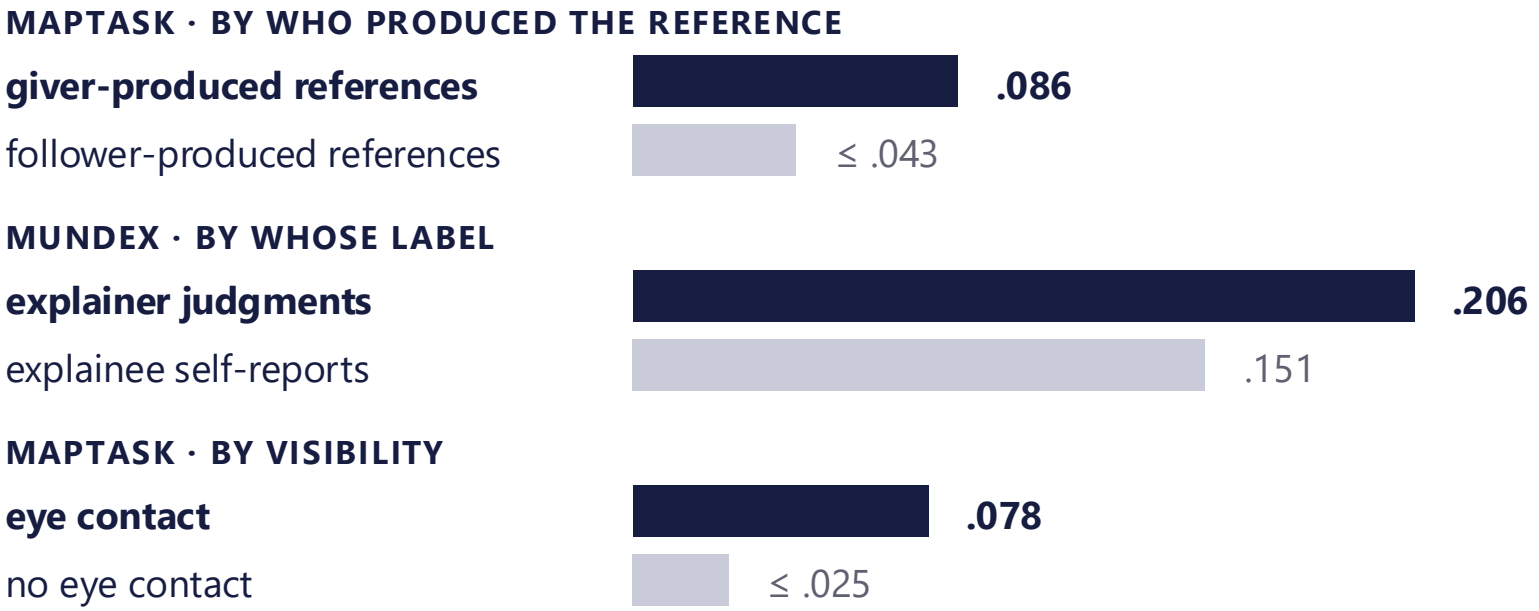
Same direction in both corpora, small effects



Four largest |r| per corpus. All Mann–Whitney p < .001; † also significant under cluster-robust GEE (FDR q < .05).

Q1 ASSOCIATIONS · BY ROLE AND VISIBILITY

Clearest in giver-produced references and explainer judgments



Largest |r| within each stratum.

The participant leading the task : givers produce the instructions being grounded; explainers monitor the explainee.

Speakers monitor addressees for understanding (Clark & Krych 2004).

Two judges of one person's understanding

EXPLAINER JUDGMENT

The explainer's estimate of **someone else's** understanding, formed from what the explainee shows, possibly including their gaze.

Attributed listener state (Buschmeier & Kopp 2018)

≠

44 / 135 linked pairs disagree

EXPLAINEE SELF-REPORT

The explainee's own account of **their** understanding, which they may not have displayed at the time.

Self-report

Same person, two judges, both labelling from the video afterwards.

Same sign on all 14 features; largest effect **.206** (explainer) vs **.151** (self-report).

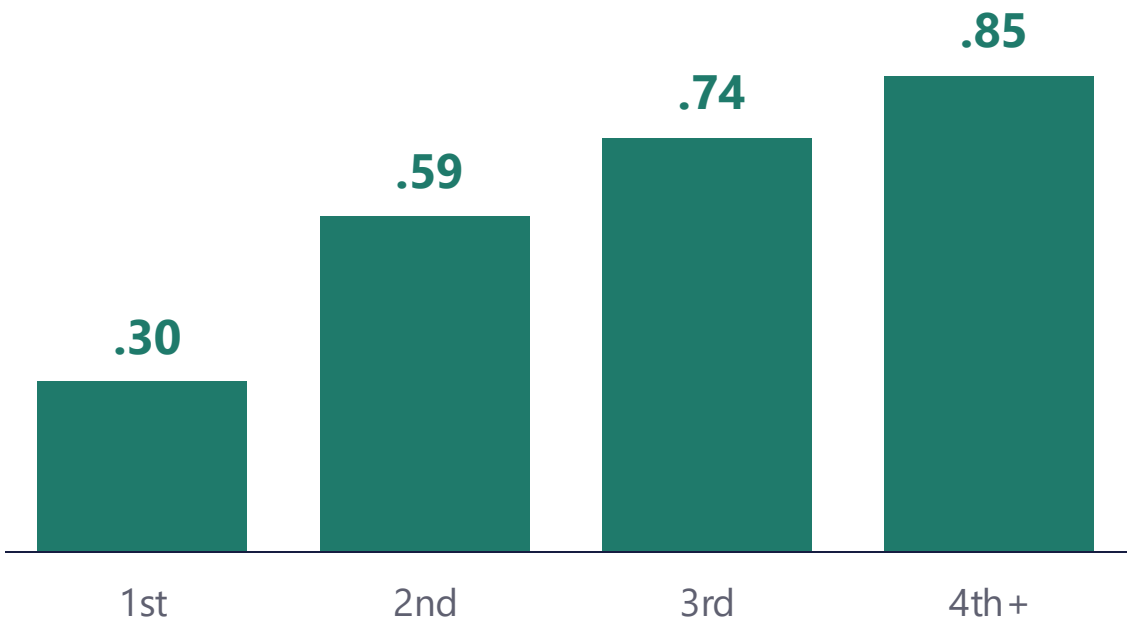
ONE POSSIBLE READING

Gaze relates most clearly to the **evidence partners use** when they judge understanding.

Q2 PROCESS · ACROSS MENTIONS

Along a reference chain, alignment climbs and gaze settles

MAPTASK · ALIGNED RATE BY MENTION



SPEAKER GAZE, 1ST → 2ND MENTION

partner gaze	.230	→	**.172**
entropy	.310	→	**.215**
transitions	.643	→	**.463**

TWO THINGS CHANGE ALONG A CHAIN

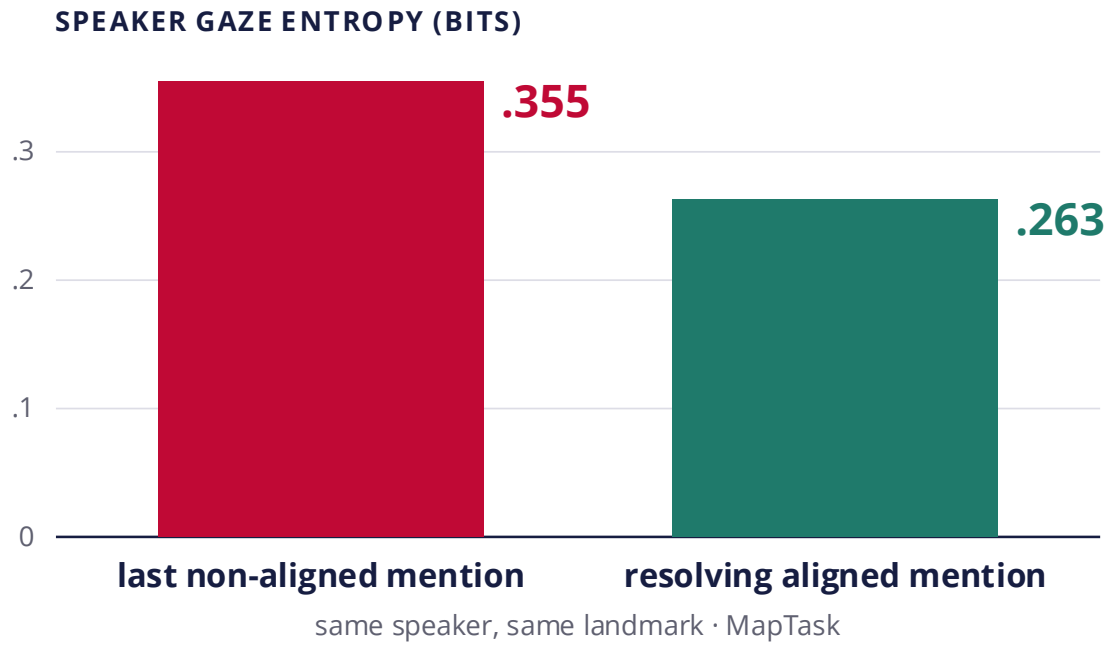
Repetition and alignment rise together.

Question : same speaker, same landmark, at the mention that resolves it.

5,144 mentions in 46 dialogues (636 · 587 · 518 · 3,403 per position). Descriptive means.

Q2 PROCESS · AT ALIGNMENT

When a referent becomes aligned, speaker entropy drops



same speaker, same landmark · MapTask

SAME DIRECTION, NOT SIGNIFICANT AFTER CORRECTION

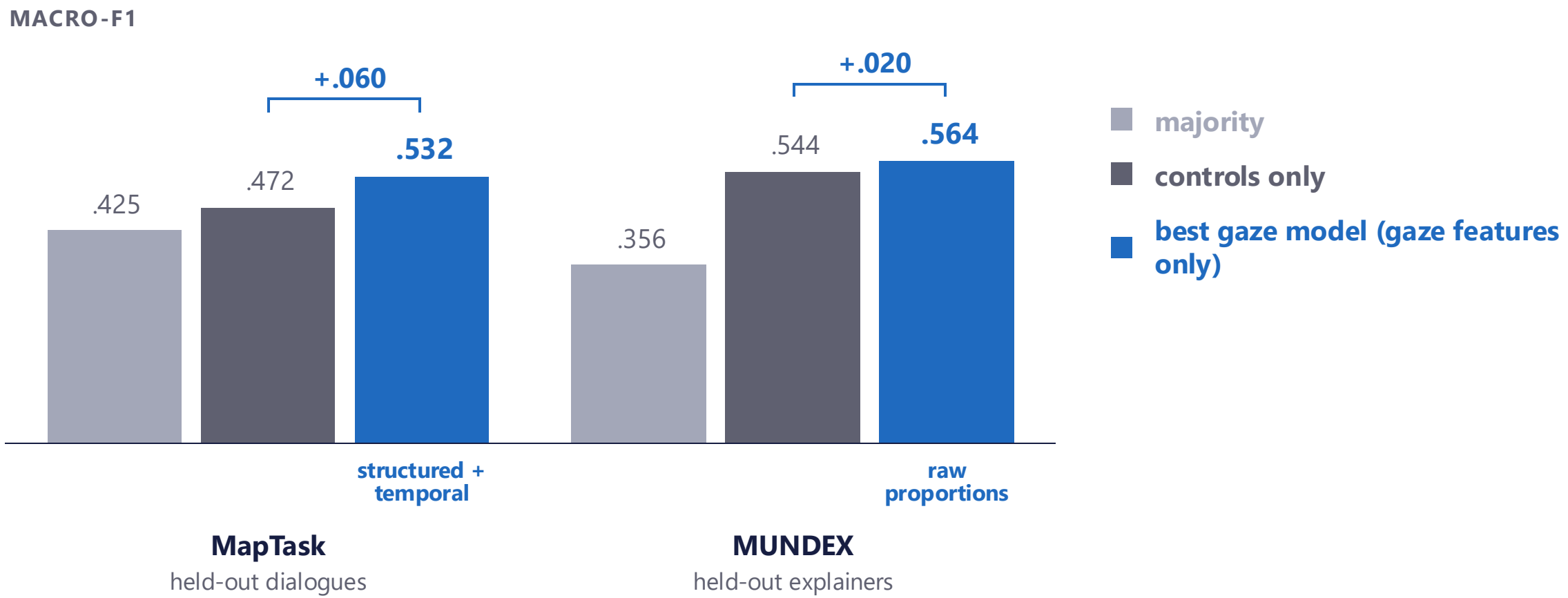
task gaze	↑
partner gaze	↓
transitions	↓

Once the referent is aligned, the speaker's gaze is less spread out.

Significant across 189 same-speaker pairs.

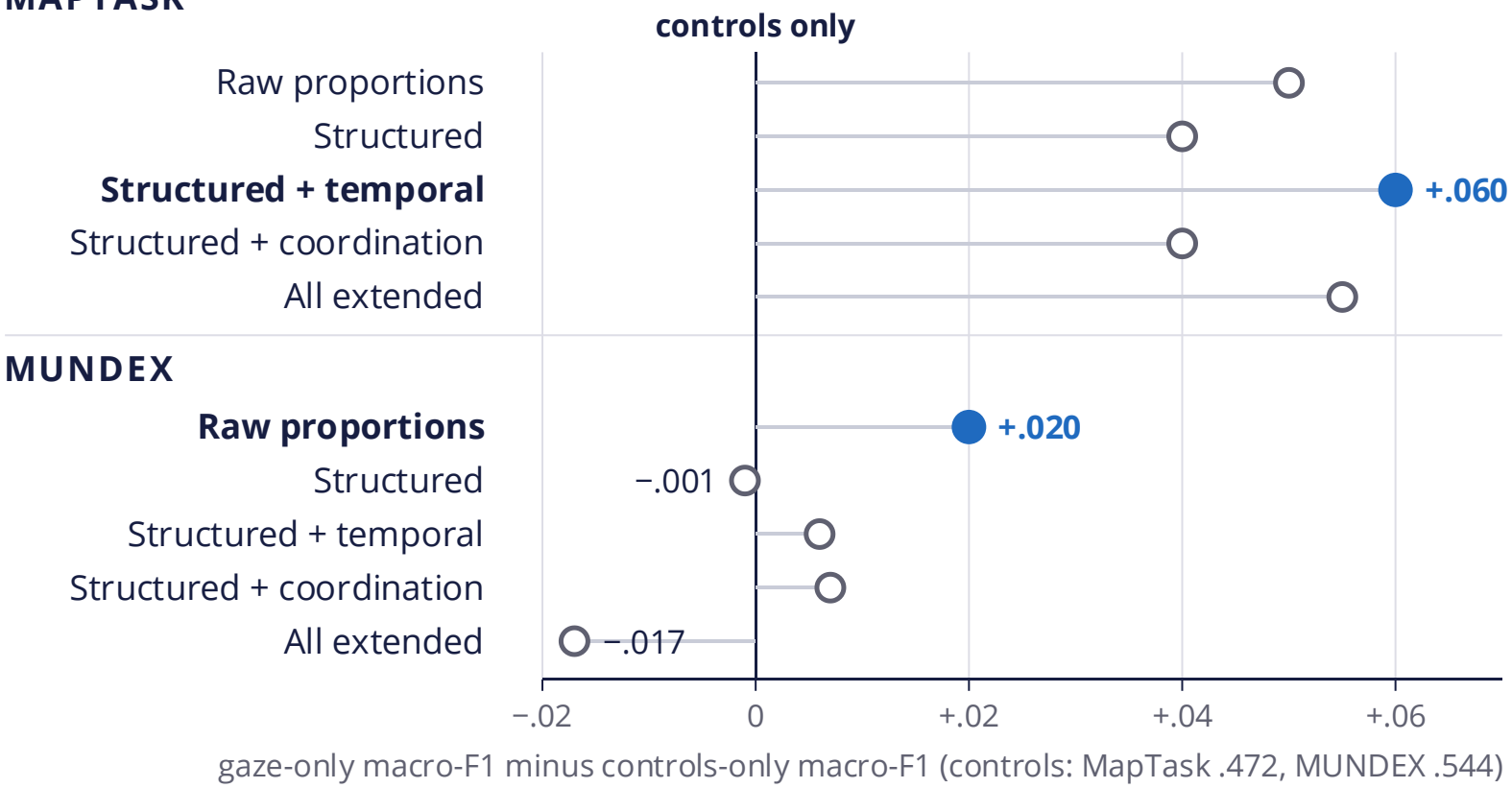
Q3 PREDICTION · VS CONTROLS

Gaze-only models beat controls-only models, modestly



Controls: task condition and speaker role (MapTask); annotator role (MUNDEX).

Timing helps in MapTask; simple proportions are best in MUNDEX



- **MapTask:** every gaze set beats controls; temporal features help most.
- **MUNDEX:** raw proportions score highest; structured and all extended do not exceed controls.

CONCLUSIONS

Where people look echoes how they ground

ATTENTION · Q1

Grounded: more task gaze, less partner gaze, lower entropy

→ **continued attention to the joint task**

MONITORING · Q1

Clearest for givers and in explainers' judgments

→ **the leading partner watches for evidence of understanding**

PROCESS · Q2

When a referent becomes aligned, speaker gaze settles

→ **gaze changes as grounding is reached**

FEATURE · Q3

Gaze alone beats controls on unseen dialogues and people

→ **a signal to combine with other cues**

Gaze is a small but consistent signal of grounding: the shared pattern spans referential alignment and reported understanding; a feature to model together with language and task context.

PAPER

arXiv:2609.18011

CODE + DATA

github.com/chnln/gaze-as-grounding-evidence

Thank you

n.li@uu.nl

REFERENCES

Works cited on the slides

Anderson et al. 1991. The HCRC Map Task Corpus. *Language and Speech*.

Bavelas, Coates & Johnson 2002. Listener responses as a collaborative process: The role of gaze. *J. Communication*.

Boyle, Anderson & Newlands 1994. The effects of visibility on dialogue and performance in a cooperative problem solving task. *Language and Speech* .

Buschmeier & Kopp 2018. Communicative listener feedback in human–agent interaction: Artificial speakers need to be attentive and adaptive. *AAMAS*.

Clark & Brennan 1991. Grounding in communication. *Perspectives on Socially Shared Cognition* .

Clark & Krych 2004. Speaking while monitoring addressees for understanding. *JML*.

Clark & Schaefer 1989. Contributing to discourse. *Cognitive Science*.

Clark & Wilkes-Gibbs 1986. Referring as a collaborative process. *Cognition*.

Doherty-Sneddon & Phelps 2005. Gaze aversion: A response to cognitive or social difficulty? *Memory & Cognition*.

Lazarov & Grimminger 2026. How are gaze aversions and mutual gaze related to the topical development of dyadic explanatory interactions? *J. Nonverbal Behavior* .

Li, Gatt & Poesio 2026. Grounded misunderstandings in asymmetric dialogue: A perspectivist annotation scheme for MapTask. *LREC*.

Murat & Vogel 2026. Gaze behaviour & conversation unfolding in the HCRC Map Task corpus. *ISA-22*.

Nakano et al. 2003. Towards a model of face-to-face grounding. *ACL*.

Türk et al. 2023. MUNDEX: A multimodal corpus for the study of the understanding of explanations. *MMSYM*.

Wang et al. 2026. Predicting states of understanding in explanatory interactions using cognitive load-related linguistic cues. *LREC*.

BACKUP · WINDOWS

Window length barely changes the results

MAPTASK

Reference + 1.5 s post-context, to catch the addressee's immediate reaction

Post-context	0 s	1 s	1.5 s	2 s	3 s
Macro-F1	.509	.520	.512	.510	.506

MUNDEX

Annotation ± 2 s : labels are short (about 1 s), so take context on both sides

Context	±1 s	±2 s	±3 s
Macro-F1	.543	.543	.549

Windows with < 30% gaze coverage dropped: **17** in MapTask, **149** in MUNDEX (139 of the 149 from two explainers).

Structured feature set, logistic regression, macro-F1; same grouped CV as the main results.

BACKUP · FEATURE GROUPS

Six feature groups, from raw proportions to all extended

Group	What it measures	Example	Features (MapTask / MUNDEX)
Raw proportions	Share of window time on each target, per person; mutual gaze	speaker task gaze	7 / 8
Structured	Raw + coverage, entropy, transitions, per person	entropy 0 = all time on one target	13 / 14
+ temporal	Gaze runs: counts, durations, switch rate, latency, first / last / dominant target	how soon the first partner look comes	+42 (21 per person)
+ bigrams	Proportions of ordered switches	share of task → partner switches	+12 (6 per person)
+ coordination	Joint state of both partners, sampled at ~10 Hz	both on task; joint entropy; partner coupling	+6
+ ratios	Relative preferences	partner/task ratio; task dominance; partner asymmetry	+9
All extended	All of the above		82 / 83

MapTask: speaker / addressee of the reference. MUNDEX: explainer / explainee, whoever gave the label. MUNDEX also has annotated mutual gaze (8th raw feature).